INFERRING DOCUMENT UTILITY VIA A DECISION-MAKING BASED RETRIEVAL MODEL

Lynda Tamine, Mohand Boughanem
Université de Toulouse
Institut de recherche en informatique de Toulouse SIG-RI
118, route de Narbonne, 31062 Toulouse (France) CEDEX 09
tamine@irit.fr, bougha@irit.fr

Keywords: Personalized information retrieval, Influence diagrams, Aggregation operators, Decision-making

Abstract: It is well known that a query is an approximate representation of the user's information needs since it does not provide a sufficient specification of the attended results. Numerous studies addressed this issue using techniques for better eliciting either document or query representations. More recent studies investigated the use of search context to better understand the user intent, driven by the query, in order to deliver personalized information results. In this article, we propose a personalized information retrieval model that leverages the information relevance by its usefulness to both the query and the user's profile, expressed by his main topics of interest. The model is based on the influence diagram formalism which is an extension of Bayesian networks dedicated to decision problems. This graphical model offers an intuitive way to represent, in the same framework, all the basic information (terms, documents, user interests) surrounding the user's information need and also, quantify their mutual influence on the relevance estimation. Experimental results demonstrate that our model was successful at eliciting user queries according to dynamic changes of the user interests.

1 INTRODUCTION

The advances in information retrieval (IR) research since the $1970's$, outline that there are two main research areas: system-oriented IR and user-oriented IR. While the former (earlier in IR research advances) focuses on document and query representation, and on techniques and models for matching such objects regardless of human environment, user-oriented research focuses on user's problem spaces: information needs formulation, information relevance statement, interactions with intermediaries, task dependency, etc.
User-oriented IR comes from the general view of cognitive IR (De Mey, 1977) that suggests that: *"... any processing of information, whether perceptual or symbolic, is mediated by a system of categories or concepts which, for the information processing device, are a model of his [its] world... "*. More precisely, there are five central dimensions of the cognitive view (De Mey, 1977):

1. information processing takes place in senders and recipients of messages,
2. processing takes place at different levels,
3. during communication of information, any actor (either sender or recipient) is influenced by its past and present experience (time) and its social, organizational and cultural environment,
4. individual actors influence the environment or domain,
5. information is situational and contextual.

According to this view, numerous critical studies (Dervin and Nilan, 1986; Shamber, 1994; Ingwersen, 1996) highlighted the limits of system-oriented IR approaches and showed the benefits of considering the information seeking environment in order to achieve more accurate interpretation of fundamental notions in IR and information seeking, such as relevance, information need and user interaction. These findings have been widely exploited in contexual IR, an active research, that has been further boosted by the increasing information on the Web and the diversity of authors and users. Indeed, a great interest has emerged recently towards the design of contextual search engines that deliver accurate results to the user according to various factors of retrieval context: interests, goals, tasks, preferences, location, time, application, etc. Numerous works in this research area consider especially the user interests, either short-term interests or long-term ones, as the main factor of the search context. Short-term interests represent the surrounding information which emerges from the current user information need in a single session. Long-term interests refer generally to the user domains of interest that have been inferred across his search history. According to Ingwersen (Ingwersen, 1996), the user interests constitute a background beyond what he typically formulates his information need and assesses the relevance and utility of the information provided by the IR system. This has turned traditional IR towards personalized IR that aims to customize information delivery according to the user profiles characterized by specific general interests.

Our contribution in this area attempts to overcome the limit of solely query-document matching by considering the user interests as clues for information relevance estimation. In attempt to achieve this goal, we propose a model which is able of making decisions in order to state about the relevance of a document w.r.t. a query depending not only on its topical relevance for the current query, but also its usefulness according to the user context, expressed through his various interests; furthermore, we address the problem of determining the most relevant context that allows to tailor the results to the intent of the current query. Thus, we turn personalized IR in a decision-making problem that can be addressed by means of an influence diagram (ID) (Shachter, 1988), which is an extension of Bayesian networks to decision-making problems. ID have a high expressive power and offer flexibility for representing and explaining decision models.
The remainder of this paper is organized as follows. In section 2, a literature of related research is presented. Section 3 presents a formal description of the problem addressed in the paper, followed by the main research objectives and motivations. Section 4 details the proposed personalized retrieval model. Section 5 presents our framework evaluation. Section 6 presents and discusses experimental results obtained using an enhanced TREC data collection. Section 7 concludes the paper.

2 Personalized vs. contextual information retrieval: an overview

Adapting a retrieval system to specific users or group of users has already been the challenge addressed by numerous Interactive Query Expansion (IQE) (Harman, 1988; Magennis and Rijsbergen, 1997) *(Tamine et al., 2003)* and Automatic Query Expansion (AQE) techniques (Mitra et al., 1998). IQE techniques like relevance feedback based strategies, employ evidence from documents, explicitly judged by the user, in order to iteratively reformulate his initial query with the aim of better fitting his information need. The main related idea is to give a part of control to the user in order to tune the formulation of the query by selecting appropriate terms or sources. AQE strategies consist in adding to the query, terms issued from the top N documents selected by the system and/or related semantic sources identified via semantic dictionaries or statistical based term selection methods.
With the increasing amount of information on the Web and the wide diversity of users, traditional AQE and IQE techniques are unlikely to work well in all retrieval situations (Ruthven, 2003; Kelly and Fu, 2007). The main problems are attributed to the extra time required to interact via the user interface, the task complexity that leads to erroneous user's suggestions for query reformulation, the inadequacy of expansion strategies to elicit ambiguous and short queries.
Another approach for adapting retrieval to specific users comes from Ingwersen works (Ingwersen, 1996) that identify various kinds of information features related to domains, tasks, interests and preferences surrounding the IR application, that can be potentially useful for contextual retrieval. Rather than exploiting short-term evidence from judged or inferred information like in IQE and AQE strategies, contextual retrieval strategies evolve a continuous

process of user adaptation through successive retrieval sessions, via the exploitation of various factors of retrieval context like judged information, interests, preferences, time, location, user's expertise etc. We attempt to elicit in the following the main concept of *context* in IR and then focus on the user's dependent context exploited in information personalization.

2.1 What about *context* in information retrieval?

The notion of *context* has a long history in multiple computer science applications (Schilit et al., 1994; Ryan et al., 1997; Davies et al., 1998). It is a wide and difficult notion that does not have one definition that can cover all the aspects it refers to. We shall basically define application context as: *"any knowledge or elementary information characterizing the surrounding application (user, objects, interactions) and having an important relationship with the application itself"*. Particularly, the context in IR applications refers back to the cognitive structures embedded in situations of retrieval or information seeking. Intuitively, user-system interaction constitutes a rich repository of potential information about preferences, experience and knowledge, as well as interests (Ingwersen and Jarvelin, 2005). This information repository represents *context of interaction*, viewed as source of evidence that could allow retrieval systems to better capture user's information needs and to more accurately measure the relevance of the delivered information. In other words, the system's estimation of the relevance would rely not only on the results of query-document matching but also on user's context-document adequacy. This has challenged the design of contextual IR systems towards the definition of relevant factors of context and the specification of methods and strategies dealing with these aspects in order to improve the search performance (Crestani and Ruthven, 2007) *(Tamine et al., 2009)*.

We outline in Table (1), five (5) context factors, listed below, that have been considered in the related literature. For each of these factors, we cite the main research works, we give the main objectives addressed by information contextualization and then present an overview of the elementary information used to build the context, and also present the involved retrieval strategy or technology.

1. *User behaviour*: user behaviour data consist in user-search engine interaction features like clickthrough data, eye-tracking, browsing features etc. These various interaction features are examined and interpreted across multiple sets of results and then used for tuning the accuracy of the delivered information. They constitute the dynamic context about the user's experience allowing to make a robust prediction about his preferences and short-term or long-term preferences, when seeking information. User behaviour is the most important factor studied these recent years in the research area of contextual IR.
2. *User interests*: generally, this factor expresses the cognitive background of the user that has an impact on his relevance judgement. The great benefit behind using the user interests, is to disambiguate queries and improve retrieval precision within a large scale of information. Numerous works in the area focussed on this aspect of context; they will be developed in section 2.2.

3. *Application*: user's application refers to the the user's background in accordance with the principles of evidence-based domain of interest like diagnosis in medicine. Domain-dependent applications provide clues allowing to better elicit the user's information need. The main objective for using application-dependent features in the retrieval process is to interpret more accurately the user's information need in a more restricted domain so to provide specific answers. In order to achieve this goal, specific domain tasks are identified with related specific queries and guidelines for selecting relevant results.

4. *Task*: task could be defined as the goal of information seeking behavior (Kelly and Belkin, 2004). Numerous tasks may be achieved by the users like reading news, searching for job, preparing course material, shopping, etc. The aim behind considering this factor is understanding the purpose of user queries in order to deliver more accurate results. In web document retrieval, user queries can achieve three main (general) tasks: the topic relevance task, the homepage finding task, and service finding task. Appropriate query and document features are then exploited in order to predict the desired task and re-rank the results. In mobile IR, task could be defined as the applications' achievement such as *tourist guide* or *GPS*[1] *based transport*.

5. *Location*: this factor concerns the geographical zone of interest corresponding to the current query. It is particularly used to categorize queries into local or global ones. Comparatively to global queries, local ones are likely to be of interest only to a searcher in the relatively narrow personal region. As an example, to look for housing is a location-sensitive query. Techniques are involved in order to identify the implicit geographical locality addressed by a query in order to improve the query results quality.

2.2 The user interests: a dominant part of context

In our contribution, we focus our efforts on the user interests as the major element of context. Both short-term and long-term interests could be embedded within his profile and are an important part of his general context, allowing better interpretation of situational relevance in contrast to topical one, based only on the query content (Borlund, 2003). They constitute a cognitive background under which user activities occur within a given retrieval session. Considering the user interests during document retrieval, leads precisely to information personalization. The key idea behind personalizing IR is then to customize search based on specific user interests. Therefore, as a personalized search engine is intended for a wide variety of users with different goals, preferences and interests, it has to learn the user model first and then to exploit it in order to tailor the retrieval task to the given user.

Numerous works in IR address the first critical question of user modelling, called also user profiling, particularly using data mining (Mobasher, 2007) or machine learning (Webb, 2001) strategies. User models or profiles consist of various and dynamic information from which appropriate techniques infer the user's background information like topics of interest,

[1]Global Positioning System

CONTEXT	*Main references*	*Main objective(s)*	*Information sources*	*Retrieval technology*
User behaviour	(Joachims et al., 2007; Agichtein et al., 2006; Shen et al., 2005; Teevan and Dumais, 2005)	Prediction of user's preferences , Inferring user's intent	Past click history, Clickthrough data, Browsing features, Query text features, Eye movements	Machine learning, Langage model over interaction features distribution
User interests	(Sieg et al., 2004; Speretta and Gauch, 2005; Liu and Yu, 2004; Ding and Patra, 2007; Jeh and Widom, 2002; Lieberman, 1995; Mobasher et al., 2000)	Query reformulation, Personalized document re-ranking	User profile, General profile issued from ontology, Judged documents, Background applications, Search history	Query-profile mapping via machine learning based algorithms, Adapted PageRank algorithm, Relevance feedback, Clustering
Application	(Lee et al., 2005; Lin and Demner-Fushman, 2005; Leuski, 2005)	Use of application dependent contextual features to adapt the results	Application procedures, data and rules	Task model
Task	(Beitzel et al., 2007; Kang and Kim, 2003; Westerveld et al., 2001; Cheverst et al., 2000)	Mapping results to query type	Document structure (Anchor, URL), Query vocabulary	Rule-based automatic classification, Machine learning
Location	(Muresan et al., 2006; Watters and Amoudi, 2003; Gravano et al., 2003; Tao et al., 2003; Yau et al., 2003; Hattori et al., 2007)	Document re-ranking considering the geographic locality of the information	Query vocabulary, URL Websites, Query history	Matching techniques, Language model, query reformulation

Table 1: A Synthetic overview on the use of context in IR

familiarity with the query topic, intent (achieved task), preferences, etc. User topics of interests are generally expressed using flat term-based vectors or vector classes (Gowan, 2003) or rich semantic structures enhanced with the use of ontologies (Liu and Yu, 2004; Speretta and Gauch, 2005; Micarelli and Sicarrone, 2004) *(Daoud et al., 2008)* extracted from various information sources: domains of expertise, logs, clickthrough data, etc.

This paper focusses on the second critical question related to the ranking model that considers the learned interests of the user (representing his profile) when computing the relevance of a document. While prior numerous works exploited the user's profile in a filtering task (Arampatzis et al., 2001; Zhang and Callan, 2001; Tebri et al., 2005; Mostafa et al., 2003), other ones (closer to our work) exploited them in a retrieval one (Sieg et al., 2004; Speretta and Gauch, 2005; Liu and Yu, 2004; Bai et al., 2007).

The related literature reveals that the user's profile is mainly exploited at the pre-retrieval step such as query reformulation, or at the post-retrieval step such as re-ranking. According to the first approach, the profiling component of *ARCH* (Sieg et al., 2004) manages a user's profile containing several topics structured as a concept hierarchy derived from assumed relevant documents using a clustering algorithm, in order to identify related semantic categories. Personalization is achieved via query reformulation based on information issued from selected and unselected semantic categories.
In (Shen et al., 2005), the authors propose several context-sensitive retrieval algorithms based on statistical language models to combine the precedent queries and clicked document summaries with the current query for better ranking the documents. More precisely, they used the clickthrough history to update the query language model and then, they compute the KL-divergence between the document language model and the updated query language model leading to the score of the document.

The second approach consists in re-ranking documents according to their closeness to the user's profile features (Speretta and Gauch, 2005; Liu and Yu, 2004; Bai et al., 2007). In (Speretta and Gauch, 2005), the authors model the user interests as weighted concept hierarchies extracted from the user's search history. Personalization is carried out by re-ranking the top documents returned to a query using a RSV[2] function that combines both similarities document-query and document-user. In (Liu and Yu, 2004) a user profile consists of a set of interests expressed as a set of semantic categories related to the query. Retrieval effectiveness is improved using voting-based merging algorithms that re-rank the documents according to the most related categories to the query. In (Bai et al., 2007) the authors integrate query context and domain context within a unified framework based on language models. Each contextual factor determines a different ranking score, and the final document ranking combines all of them.
Although the motivation is similar, our approach is different from the cited previously works in two main points:

1. our approach for personalizing document ranking is performed at the retrieval stage as it exploits the user's profile as an explicit part of the formal ranking model, and not as a source of evidence to reformulate the query or re-rank the documents;
2. we turn the information personalization problem to a decision-making task. For this aim, we explore the use of ID which are Bayesian probabilistic tools dedicated to decision-making problems.

Our goal is to show how user interests (either short-term or long-term ones) could be explicitly integrated into a unified model and combined in order to evaluate the global utility of the decisions related to the relevance of documents w.r.t. a query, considering the surronding context. Our contribution in this paper focusses on the personalized retrieval model. We assume that the user's profile is described by a set of general interests covering his cognitive background. Each user interest expresses a specific topic that has already emerged across the

[2]Relevance Status Value

user's search history. The user profile is learned using an appropriate methodology (which is not the focus of our study) based on implicit relevance estimation of the information gathered through successive retrieval sessions *(Daoud et al., 2008; Daoud et al., 2009)*. Each user interest is represented using a term-weighted vector where each term represents a dominant keyword. This offers flexibility to plug our personalized retrieval model to various user modelling strategies.

3 Problem specification and motivations

3.1 Defining information personalization as a decision-making problem

A decision-making problem is typically defined through a preference relation established among a set of decisions that aim to retain alternatives a_i among the set of all possible ones A according to a utility function μ. A preference relation between alternatives a_i and a_j, denoted $a_i \succ a_j$, allows to support the decision "*alternative a_i is retained rather than alternative a_j*". In this paper, we address the problem of personalizing IR as a decision-making problem under uncertainty as follows:
Given a query Q, the IR problem is to rank a set of documents D according to their relevance to the information need of the user profile U represented using a set of topics of interest such as $U = \{C_1, \ldots, C_p\}$, where p is the number of interests in the user profile. We consider that the computation of the relevance measure involves choosing some actions among a multitude of actions: decide which interest(s) covers the query topic, which document is relevant to the query, which document is relevant to any interest, etc. Thus, there are several criteria for measuring the relevance of a document w.r.t. the query topic and the user interests. More precisely, we shall combine two main levels of relevance:

1. *over a user interest*: at this first level, we gauge how much a document is relevant or covers the query topic with respect to each of the user interests represented in his profile, separately. Our goal is to measure the intrinsic relevance of a document $D_i \in D$ according to a specific user interest C_k belonging to the user profile U. More precisely we shall consider different sources of evidence related to the content of the document and the content of the given user interest. This leads to a preference relation, noted below $\succ_C$, that ranks the documents w.r.t. each user interest.
2. *over the user's profile*: at this second level, we gauge how much a document answers the query topic by covering as much as possible all the user interests representing his profile. We consider that the user interests represent criteria of relevance of the documents to be presented to the user. This leads to combine at this level, the ranking involved at the previous level of relevance, in order to state the final ranking of documents using the preference relation, noted below $\succ_U$.

With this in mind, we define preference relations at each level of relevance as follows:

- ***Preference relation at the user interest level***: Given a pair of documents $(D_1, D_2) \in D \times D$, relation $\succ_C$ is a preference relation over the user interest (context) C so that:

$$(D_1 \succ_C \ D_2) \equiv (D_1 \ is \ preferred \ to \ D_2 \ according \ to \ the \ user \ interest \ C)$$

- ***Preference relation at the user profile level***: Given a pair of documents $(D_1, D_2) \in D \times D$, a relation $\succ_U$ is a preference relation over the user's profile U so that:

$$(D_1 \succ_U \ D_2) \equiv (D_1 \ is \ preferred \ to \ D_2 \ according \ to \ the \ user \ profile \ U)$$

Typical properties of the relations $\succ_C$ and $\succ_U$ are assymetry and transitivity.

From the probabilistic point of view, the preference relations introduced above are induced by the ranking fuction of the documents D considering the query Q and the user profile U, noted $p(d|q,u)$. We define the preference relation $\succ_U$ as:

$$D_1 \succ_U D_2 \Leftrightarrow p(d_1|q,u) > p(d_2|q,u) \tag{1}$$

where d_i, q and u are the random variables associated to respectively document D_i, query Q and user profile U and:

$$p(d_i|q,u) = \frac{p(q|d_i,u)p(d_i|u)}{p(q|u)} \tag{2}$$

As the denominator $p(q|u)$ is independent from d for a given query and user, we can use only the numerator in order to rank the documents. Thus, we define the relevance of document D_i according to the query Q and the user profile U, noted below $RSV_U(Q, D_i)$, as:

$$RSV_U(Q, D_i) = p(q|d_i,u)p(d_i|u) \tag{3}$$

The first term of equation (3) is query dependent reflecting the closeness of document D_i and query Q according to user U. The second term is query independent, highlighting the usefulness of the document to the user. This may express the suitability of the document to all the domains of interest of the user when seeking information. Assuming that the user profile is modelled using a set of interests such as $U = \{C_1, C_2, ..., C_p\}$, the formula (3) gives:

$$RSV_U(Q, D_i) = p(q|d_i, c_1, c_2, ..., c_p)p(d_i|c_1, c_2, ..., c_p) \tag{4}$$

where c_k refers to a random variable associated to the user interest C_k. The formula (4) highlights that:

1. two key conditions are prevalent when computing the relevance of documents:
   (a) relevance condition, expressed by $p(q|d_i, c_1, c_2, ..., c_p)$, that ensures that the selected documents are close to the query Q;
   (b) the usefulness condition, expressed by $p(d_i|c_1, c_2, ..., c_p)$ that ensures that the selected documents are consistent with the user interests $\{C_1, C_2, ..., C_p\}$;
2. we assume that maximum likelihood of a document is achieved when maximizing the coverage of the information according to the different user interests. The user may choose the degree of relevance to integrate either all, or a sublist of main topics of interest during the personalization process.

By considering this manner of addressing the problem of personalized IR in the context of user's multi-interests, we are hence attracted by formulating it in a mathematical model based on a utility theory supported by ID wich are extension of Bayesian models. The problem is globally expressed through $ID(dn, cp, rn, \mu, \Psi)$:

- document variable set $dn = \{d_1, d_2, ..., d_n\}$ where n is the number of documents in the collection,
- user interests variable set $cp = \left\{c_1, c_2, ..., c_p\right\}$
- decision variable set $rn = \{r_1, r_2, ..., r_n\}$ where r_i is the decision of stating that document D_i is relevant,
- utility function attached to each decision r_i; it quantifies the quality of the decision r_i. Thus, we attach to each document D_i a set of utility values $\mu^i = \left\{\mu_1^i, ..., \mu_p^i\right\}$ where μ_k^i expresses the utility of the positive decision r_i about the relevance of a document D_i according to the user interest C_k (that supports the preference relation $\succ_C$), noted below $\mu(r_i|d_i, c_k)$.
- Ψ is an aggregation operator expressing the joint utility that combines evidence values from the whole user profile $U = \left\{C_1, C_2, ..., C_p\right\}$

With respect to the probabilistic view illustrated above, the problem of personalized IR takes then the form of ranking the documents $D_i \in D$ according to:

$$RSV_U(Q, D_i) = \Psi_{k=1..p}(\mu(r_i|d_i, c_k)p(q|d_i, c_k)) \tag{5}$$

Section 4 gives formal details of our personalized IR model based on the above specification.

3.2 Main research objectives

The current state of (out-of-context) IR research spans by the two dimensions: *Domain* and *Research & Development* areas (Jose and Rijsbergen, 2005). Our research has been carried out in the design of IR models within the above dimensions and a third one: *context*. The main related goal is to achieve the specification of domain-dependent IR models that can be used to focus an IR task according to the user's domain(s) of expertise like medicine, library, e-commerce, legal IR etc. For this, both IR and context models have to be reexamined in order to personalize IR.
Our objective in this paper is to highlight the prevalence and the usefuleness of the evidence extracted from multiple user interests, embedded within his general expertise, in order to tune the accuracy of the results presented in response to the query. We particularly explore two major research questions in this study:

1. How to model a personalized retrieval task within a broad variety of topics of interest? As previously discussed, the user's searches may have multiple goals or topics of interest and occur within the broader context of their information-seeking behaviors. The problem we address is to infer the information utility according to the query and the user's cognitive background, expressed using a set of various interests.

2. How to combine evidence from these various interests (belonging to either related or unrelated topics) in order to measure the relevance of a document, in response to a specific user query? We explore whether several aggregation operators applied to the information utility, can be successfully exploited in order to improve the search performance.

4 A decision theoretical model for personalized information retrieval

In this section we detail our personalized retrieval model. We start by presenting the background required to build our model and then we proceed to the description of the corresponding qualitative and quantitative components.

4.1 Background

4.1.1 Influence diagrams: Bayesian network extension

A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest (Jensen, 2001). A Bayesian network uses qualitative and quantitative components to model and manipulate n-dimensional probability distributions. The qualitative component is carried out through a Directed Acyclic Graph (DAG), $G =< V, E >$ where each node in $X_i \in V$ encodes the random variable of interest and E encodes the relationships among these variables. We note $Pa(X_i)$ the parent set of X_i in G. The quantitative component outlines the estimation of the conditional dependencies among the variables. More precisely, for each variable $X_i \in V$, is attached conditional probability distributions $p(X_i|pa(X_i))$ where $pa(X_i)$ represents any combination of the values of the variables in $Pa(X_i)$. The inference of new sources of evidence is done using the joint distribution law:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} (p(X_i|pa(X_i)) \tag{6}$$

An ID (Shachter, 1988) is a Bayesian probabilistic model extension devoted for solving decision-making problems. The basis of an ID are probabilities and utilities. Utilities are quantified measures for preference, attached to each possible situation (scenario) concerned by the modelled decision-making problem. In practice, an ID is represented by an acyclic DAG containing qualitative and quantitative components.

- *Qualitative component*: there are three types of nodes (chance, decision and utility nodes) and two main types of arcs (informative and influence arcs). *Chance nodes*, denoted X, are usually drawn as circles; they are represented using random variables x that are relevant to the decision problem and cannot be controlled. A configuration of instantiated chance nodes $\{x_1, \ldots, x_l\}$ expresses a set of observations related to a specific situation s_i among all the possible ones $S = \{s_1, \ldots, s_k\}$ that can occur within the modelled decision problem. *Decision nodes*, usually drawn as rectangles, represent variables that the decision maker controls directly. *Utility nodes*, usually drawn as diamons, express the preference degree of the benefit attached to the consequences derived from the decision process. They are quantified by the utility of each possible combination of their parent nodes.

There are also different kinds of arcs that join the ID nodes. The arcs between chance nodes, or between a chance node and a decision node are called informative arcs. The arcs between chance nodes express probabilistic dependencies as in Bayesian networks. The arcs from a decision node to a chance node or to a utility node, express the fact that the incoming decision will influence the value of the chance node or the profit obtained. The influence arcs join chance nodes to utility nodes and express the fact that the benefit depends on the value that these chance nodes take.

- *Quantitative component:* the dependencies between chance nodes, representing random variables, are carried out using classical Bayesian probability distributions. Practically, for each chance node X in the graph, is attached a set of conditional probability distributions $p(x|pa(x))$, one for each configuration $pa(x)$ from the parent set of the node X in the graph. For each utility node V, related to a decision node, is attached a random variable v and a real-valued function μ over its parents $Pa(V)$ specifying for each situation $s \in S$, expressed by a configuration of the instance values $pa(v)$ ($s = pa(v)$), a measure denoted $\mu(pa(v))$. This measure quantifies the benefit attached to this configuration (viewed as a set of observations) for the decision maker.

Given a particular situation $s_i \in S$, the diagram evaluation is carried out using an evidence propagation algorithm which aims to determine the decision alternative that will lead to the optimal utility called *expected utility*, denoted $EU(s_i)$. Several aggregation operators can be used in order to compute the joint expected utility. The following paragraph gives a brief overview of their formal properties.

4.1.2 Aggregation operators

Aggregation of information is the simultaneous use of different pieces of information provided by several sources, to come to a conclusion or a decision (choice, rank etc.). Aggregation operators are mathematical functions, which assign a utility value for alternatives gathering of different values according to different criteria. Each criterion is a factor that has an impact on the quality of the decision.
More formally, an aggregation operator, defined on a set of alternatives $A \in R^n$ expressed through the n criteria, $cr_1, \ldots, cr_n$, is expressed as follows:

$$\begin{aligned} R^n &\rightarrow R \\ (a_1,\ldots,a_n) &\longmapsto \psi(a_1,\ldots,a_n) \end{aligned} \tag{7}$$

where $(a_1,\ldots,a_n)$ is the descriptor of the alternative $a \in A$, a_i represents the measurement of the criterion cr_i. An aggregation operator satisfies the main following properties:

- *Idempotence*: $\psi(X) = X$
- *Boundary conditions*: $\psi(0,\ldots,0) = 0 \;\; \psi(1,\ldots,1) = 1$
- *Non decreasing*: $\psi(x_1,\ldots,x_n) \leq \psi(y_1,\ldots,y_n) \; iif \;\; (x_1,\ldots,x_n) \leq (y_1,\ldots,y_n)$

Various classes of aggregation operators have been proposed (Fargier and Perny, 2003): Min and Max, Weighted Minimum and Maximum, Ordered Average Operators .

An aggregation operator allows ranking of alternatives according to a preference relation defined as follows:

$$a \succ b \equiv \phi(\psi(a), \psi(b)) \tag{8}$$

where ϕ is an ordered function taking values in $\{0,1\}$. The usual expression of the function ϕ is:

$$\phi(x,y) = \begin{cases} 1\ if\ y - x \leq p \\ 0\ otherwise \end{cases} \tag{9}$$

where p is a threshold discrimination value.

4.2 The personalized retrieval model

Our research considers the following retrieval setting: a user U interacts with a document space $D = \{D_1, D_2, \ldots, D_n\}$ with a typical search engine through a query Q. D_i is the i^{th} document represented as a term vector using the index $T = \{T_1, T_2, \ldots, T_m\}$. In this setting, users have *a priori* general topics of interest such as $U = \{C_1, C_2, \ldots, C_p\}$. C_k is the k^{th} user interest previously learned across the search history and represented also as a flat term vector using the index T. Let $\tau(D_i)$, $\tau(C_k)$ and $\tau(Q)$ be the index terms belonging to document D_i, user interest C_k and query Q respectively.

4.2.1 Diagram topology

Figure 1 illustrates the qualitative component of our ID based retrieval model *(Tamine-Lechani et al., 2007; Zemirli et al., 2007)*. The set of chance nodes I is composed of four different types of nodes $I = Q \cup D \cup T \cup U$ expressed above. The set of decision nodes $R = \{R_1, R_2, .., R_n\}$ represents the decisions to state that these documents are relevant.
There is a utility node corresponding to each decision node. Furthermore, there is a node representing the cumulative joint utility of the model. Chance nodes and decision nodes are detailed below:

- *Chance nodes*. There are four types of chance nodes: query, documents, terms, and user interests.

  Each document node D_i, represents a binary random variable taking values in the set $dom(D_i) = \{d, \overline{d}\}$, where d expresses, as in the Turtle model (Turtle and Croft, 1990), that the document D_i has been observed and so introduces evidence in the diagram, all the remaining document nodes are set to $\overline{d}$ alternatively to compute the *a posteriori* relevance. Similarly, C_k represents a binary random variable taking values in the set $dom(C_k) = \{c, \overline{c}\}$, where c represents '*the user interest c is under consideration (observed)*' and $\overline{c}$ represents that '*the user interest c is not under consideration (not observed)*'. Each term node T_j represents a binary random variable taking values in the set $dom(T_j) = \{t, \overline{t}\}$, where t represents '*term T_j is relevant for a given query* ' and $\overline{t}$ represents that *'term T_j is not relevant for a given query'*. The relevance of a term expresses its adequacy to deal with the document topic; therefore, a term could be relevant to a document even if it does not index this

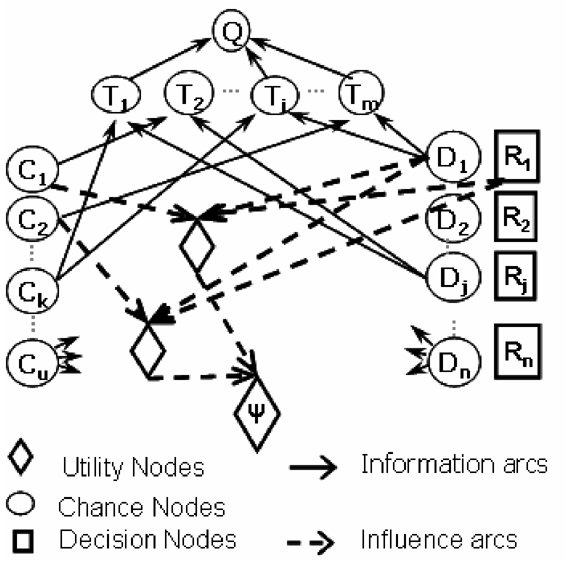

Figure 1: The diagram topology

document. In the domain value of the query $\{q, \overline{q}\}$, q means that the query is satisfied and $\overline{q}$ that it is not satisfied. As only the positive query instantiation is of interest, we consider $Q = q$ only.

- *Decision nodes*. Our ID contains one type of decision node. Each decision node R represents a binary random variable taking values in the set $\{r, \overline{r}\}$. These values represent the decision random variables related to whether the document is to be estimated as relevant and so presented to the user, or estimated as not relevant and so not to be presented.

The relationships between the nodes described above are represented by the following arcs:

- *Informative arcs*. These arcs join the chance nodes. More precisely, informative arcs join the query node Q and each related term $T_j \in \tau(Q)$. There are also arcs joining each term node $T_j \in \tau(D_i)$ and each document node $D_i \in D$. Similarly, there are arcs joining each node term $T_j \in \tau(C_k)$ and each user interest node $C_k \in U$.
- *Influence arcs*. These arcs point to utility nodes. In our current problem, the utility function depends obviously on the decision to be made, the content of the document and the adequacy of the user interest to the current query context. The aggregated utility of the model will depend on the individual utility values computed at each corresponding node.

4.2.2 Query evaluation

Query evaluation consists in the propagation of new evidence through the diagram, as in Bayesian networks, in order to maximize a ranking utility measure. In our approach, this measure is based on the global joint utility value corresponding to the most accurate decisions related to the relevance of a document according to the query and the user's profile. As specified above, this leads to rank documents according to two levels of relevance: user interest level and profile level, as detailed below.

1. **Ranking documents over one user interest**
   In practice, given a query Q represented by a set of positive terms ($\tau(Q) = \{T_1, T_2, ..., T_r\}$), the retrieval process starts by placing the evidence alternatively in each observed document node then, the inference process is run as in a decision-making problem by maximizing a re-ranking utility measure $RSV_{C_k}(Q, D_i)$:
$$RSV_{C_k}(Q, D_i) = \mu(r_i|d_i, c_k) * p(q|d_i, c_k) \tag{10}$$
   where the utility value $\mu(r_i|d_i, c_k)$ expresses the degree of the closeness between the document D_i and the user interest C_k.

- **Computing** $\mu(r_i|d_i, c_k)$. The topical closeness between a document and a user interest could be computed using various similarity measures, we propose the following one:
$$\mu(r_i|d_i, c_k) = 1 + log(1 + \frac{\sum_{T_j \in (\tau(D_i) \cap \tau(C_k))} nidf(T_j)}{\sum_{T_j \in \tau(D_i)} nidf(T_j)}) \tag{11}$$
  where $nidf(T_j)$ is the normalized idf of term T_j such as $nidf(T_j) = \frac{idf(T_j)}{Max_{T_l \in T}(idf(T_l))}$, $idf(T_j) = log(\frac{n}{n_j})$ where n is the total number of documents in the collection, n_j is the number of documents indexed with term T_j.
- **Computing** $p(q|d_i, c_k)$. This factor is computed as $\frac{p(q, d_i, c_k)}{p(d_i, c_k)}$. By applying the probability marginalisation principle and by considering the topology of our model, $p(q, d_i, c_k) = \sum_{\theta^s \in \theta} p(q, \theta^s, d_i, c_k)$ where θ represents all the possible configurations of terms in $pa(q)$, θ^s the s order configuration. By applying the joint probability of all the nodes in the graph illustrated in figure 1, we obtain:
$$p(q, d_i, c_k) = p(d_i) * p(c_k) * \sum_{\theta^s \in \theta} p(q|\theta^s) * p(\theta^s|d_i, c_k) \tag{12}$$
  Because of the important amount of calculations required on retrieval time, we propose to use an approximation by assuming that terms are independent given the documents; thus we can rewrite the formula (12)
$$p(q, d_i, c_k) = p(d_i) * p(c_k) * \sum_{\theta^s \in \theta} (p(q|\theta^s) * \prod_{T_j \in \tau(Q) \cap (\tau(D_i) \cup \tau(C_k))} p(\theta^s_j|d_i, c_k)) \tag{13}$$
  where θ^s_j the s order configuration for term T_j in $pa(q)$. For instance if Q node is related to term nodes $\{t_1, t_2\}$, $\theta = \{\{t_1 t_2\}\ \{t_1 \overline{t_2}\}\ \{\overline{t_1} t_2\}\ \{\overline{t_1}\, \overline{t_2}\}\}$, the instance θ^1_1 of T_1 in the first configuration $\theta_1 = \{t_1 t_2\}$ is $\theta^1_1 = t_1$.
  Finally, we obtain
$$p(q|d_i, c_k) = \sum_{\theta^s \in \theta} (p(q|\theta^s) * \prod_{T_j \in \tau(Q) \cap (\tau(D_i) \cup \tau(C_k))} p(\theta^s_j|d_i, c_k) \tag{14}$$

   Assuming that the user interests and the collection documents are also independent given a user, the combination of formula (10) and (14) leads to:
$$RSV_{C_k}(Q, D_i) = \mu(r_i|d_i, c_k) * \sum_{\theta^s \in \theta} (p(q|\theta^s) * \prod_{T_j \in \tau(Q) \cap (\tau(D_i) \cup \tau(C_k))} p(\theta^s_j|d_i) * p(\theta^s_j|c_k)) \tag{15}$$

Therefore, one can express:

$$\succ_{C_k}(D_1,D_2)=\begin{cases}1\ if\ (RSV_{C_k}(Q,D_1)-RSV_{C_k}(Q,D_2))\geq 0\\ 0\ otherwise\end{cases} \tag{16}$$

Considering as mentioned above, that θ^s is a possible configuration of $pa(q)$ and that θ^s_j is a possible instance of term T_j (taking values t_j or $\overline{t_j}$), we give in the following, the general formulation of the conditional probabilities $p(q|\theta^s)$, $p(t_j|d_i)$ and $p(t_j|c_k)$.

- **Computing** $p(q|\theta^s)$. As previously mentioned, the query is a leaf node that has as many parents as terms are belonging to its representation, noted by $Pa(Q)$. Therefore, it should store 2^k configurations $pa(q)$, k being the number of parents. Taking into account only the positive configuration term parents, noted $R(pa(q))$, we can compute the probability function attached to a query node using the *Noisy-Or* aggregation operator (Pearl, 1988). An important advantage of this operator is that it focusses on the positive instanciation of all the query terms; furthermore the absence of a highly weigthed term in the configuration term parent of the query is significantly penalized. According to this principle, the probability $p(q|pa(q))$ is computed as follows:

$$p(q|pa(q))=\begin{cases}0\ if\ (pa(Q)\cap R(pa(q))=\emptyset\\ \frac{1-\prod_{T_j\ \in R(pa(Q))} nidf(T_j)}{1-\prod_{T_j\ \in Pa(Q)} nidf(T_j)}\ otherwise\end{cases} \tag{17}$$

  Formula (17) means that more query terms among $Pa(Q)$ are instanciated as positive, the higher is the probability to satisfy the query Q. If $R(pa(q))$ contains all the query terms then $p(q|pa(q))=1$.
- **Computing** $p(t_j|d_i)$ **and** $p(t_j|c_k)$**.** The estimation of the conditional probabilities of relevance of a term is based on its discrimination power in the document and user interest descriptors as follows:

$$p(t_j|d_i)=\begin{cases}\frac{wtd(T_j,D_i)}{\sum_{T_l\in\tau(D_i)} wtd(T_l,D_i)}\ if\ T_j\in\tau(D_i)\\ \delta_d\ otherwise\end{cases} \tag{18}$$

$$p(t_j|c_k)=\begin{cases}\frac{wtc(T_j,C_k)}{\sum_{T_l\in\tau(C_k)} wtc(T_l,C_k)}\ if\ T_j\in\tau(C_k)\\ \delta_c\ otherwise\end{cases} \tag{19}$$

  where $wtd(T_l,D_i)$ and $wtc(T_l,C_k)$ are respectively the weights of term T_l in document D_i and user interest C_k, δ_d and δ_c constant values ($0\leq\delta_d,\delta_c\leq 1$) expressing the default probability value that represent the ignorance about the relevance of a term that does not belong to the document or user interest index; we assume that these probabilities are identical for all the term nodes, the parameters used in this paper are $\delta_d=0.5$ and $\delta_c=0.5$.
  Consequently, we compute the probability of irrelevance of a term to a document or a user interest as respectively: $p(\overline{t_j}|d_i)=1-p(t_j|d_i)$, $p(\overline{t_j}|c_k)=1-p(t_j|c_k)$.

2. **Ranking documents over the user profile**
   The problem addressed at this level concerns the joint utility estimation of a document according to the whole user profile. Usually, documents are ranked according to their relevance w.r.t. a given query Q, denoted by $RSV(Q,D)$. Various mapping functions have been proposed in the litterature (Nottelman and Fuhr, 2003) in order to approximate the relationship between RSV and probabilities of relevance. Following the decision theoretical support of our approach, the personalized RSV_U measures the accuracy of the decisions related to the relevance of the documents to be presented according to the query Q and the user profile U. We propose the following mapping function:

$$RSV_U : \begin{cases} R \longrightarrow \; R \\ RSV_U(Q,D) \mapsto EU(r|d) \end{cases} \tag{20}$$

   where $EU(r|d)$ is the expected utility of the decision *"D is relevant, to be presented to the user"*. Analogously to the preference relation $\succ_C$, the preference relation $\succ_U$ is expressed as follows:

$$\succ_U (D_1,D_2) = \begin{cases} 1 \; if \; EU(r|d_1) - EU(r|d_2) \geq 0 \\ 0 \; otherwise \end{cases} \tag{21}$$

   The global relevance measure of a document (used as the basis of information relevance) is computed with respect to the partial relevance estimations computed above within each user interest. We formally express:

$$EU(r|d_i) = \Psi(RSV_{C_1}(Q,D_i),\ldots,RSV_{C_p}(Q,D_i)) \tag{22}$$

   where Ψ is an aggregation operator. Assuming that a query covers one topic, our purpose is to determine the most suitable context to the current query by means of dependence vs. independence of the user interests. This dependence refers to the relatedness or semantic closeness of the general topics they belong to.

   - ***Hypothesis 1***: *User interests are independent*
     In this case, the rank of a document should be high according to the suitable user interest and low according to the others. This leads to apply the principle of relative agreement in decision-making (Fargier and Perny, 2003). This principle means, in our case study, that the global relevance of a document depends on the degree of mismatch of rankings across the different topics of interest. A possible formulation of the aggregation operator is:

$$\Psi_{k=1\ldots p}(RSV_{C_k}(Q,D_i)) = \left\{ \begin{array}{l} 1 \; if \; \sum_{k=1..p}(RSV_{C_k}(Q,D_i) \geq \sum_{k=1..p}(RSV_{C_k}(Q,D_j)), \forall j \neq i \\ 0 \; otherwise \end{array} \right\} \tag{23}$$

     where p is the number of user interests

   - ***Hypothesis 2***: *User interests are dependent*
     The dependence of the user interests implies a possible reinforcement of the information

relevance according to the query. This could express in some cases, the presence of subtopics of a general topic as in hierarchical representations. In this case, a suitable aggregation operator should be based on the principle of absolute agreement (Fargier and Perny, 2003) of relevance across the different topics of interest. A possible formulation of the aggregation operator is:

$$\Psi_{k=1...p}(RSV_{C_k}(Q,D_i)) = \left\{ \begin{array}{l} 1\ if\ \sum_{k=1..p}(RSV_{C_k}(Q,D_i)) > \alpha \\ 0\ otherwise \end{array} \right\} \quad (24)$$

where α is a threshold value.

5 Experimental evaluation

This section reports our experimental framework and evaluation results to validate the effectiveness of our proposed model.

5.1 Framework evaluation

It is well known that the Cranfield model (Cleverdon, 1967) is the dominant evaluation model in IR. It constitutes a laboratory model based on the availability of a document collection, a set of well defined topical queries and a set of relevance assessments identifying the documents that are topically relevant to each query. Recall, precision, and mean average precision (MAP) are generally used to express the effectiveness of IR models and algorithms within such test collections. The TREC evaluation protocol is defined according to the Cranfield model; it provides various and appropriate document collections and query sets for specific tasks enabling accurate comparative evaluation in IR. However, this evaluation approach has been challenged by the emergence of contextual IR because of the lack of resources expressing the user's context determined by various features eg. user's task, user interests, dynamic relevance, user's expertise etc. (Ingwersen and Jarvelin, 2005). For this reason alternative approaches to evaluation of contextual IR are required. The first one consists in building specific test collections and user's judgments by conducting the evaluation with real users (Teevan and Dumais, 2005; Liu and Yu, 2004). The main advantage of such evaluation is that it is realistic; however, comparative evaluation is critical.

The second alternative we applied in our experimental study, as done in previous works (Bai et al., 2007; Shen et al., 2005), aims at exploiting TREC resources and enhancing them by hypothetic users, from whom user profiles are simulated. Thus, experiments reported in this paper outline the performance of our retrieval model, taking predefined user profiles as a starting point. User interests are simulated, in the first set of experiments, using a manual annotation of general topics of interest (domains) assigned by human assessors. For each general topic of interest is attached hypothetic users with specific related user interests (sub-topics) built using different relevance judgements viewed as answers that could be those that the user have read, browsed or judged explicitly as relevant. Furthermore the simulation method we used, is based on a cross-validation strategy that could be considered as a simulation of users' changing interests, as both the training set and the test set change. Although subjective, this

approach allows meaningful observations, and testing the soundness and technical validity of the proposed model.
In order to evaluate the effectiveness of our model, we need the following three datasets:

1. *query topics and relevant judgments*: we used particularly the queries among $q_{51} - q_{150}$ because they are tagged with the domain meta data that gives the query domain of interest; below an example of topic:
   <*num*> *Number : 59*
   <*dom*> *Domain: Environment*
   <*title*> *Topic: Weather related fatalities*

   We used for our tests, the topics addressing 8 domains of interests with different sizes as illustrated in figure 2. The availability of specified query domain allows us to build simulated user interests and then to compare our personalized retrieval model to a basic one that does not take into account the user interests, or another personalized one that involves a different retrieval model/strategy.

2. *a document collection*: we used a *TREC* data set from Disks 1-3 provided for the ad hoc and filtering tracks, containing documents issued from Newspapers *Associate Press (AP)* and *Wall Street Journal (WJS), Financial Times (FT)* etc. We have choosen this collection because the query topics $q_{51} - q_{150}$, described above, are assessed on this collection.

3. *user interests*: in order to map the query domains to realistic and dynamic user interests, we used the cross-validation strategy as illustrated in the algorithm presented below. Each query set, containing n queries related to a specific domain of interest, is divided into subsets having approximately the same number of queries. We repeated experiments n times, each time using a different subset as the test set and the remaining as the training set from which we built the user simulated interests using the OKAPI algorithm. Each user interest is represented as a term based vector where the term weighting formula is: $wtc(j,k) = \log \frac{(r_j+0.5)/(r-r_j+0.5)}{(n_j-r_j+0.5)/(n-n_j-r-r_j+0.5)}$, where r is the number of relevant documents to each query in the training set belonging to C_k, r_j the number of relevant documents containing term T_j, n_j the number of documents containing term T_j, n is the total number of documents in the collection.

   We outline that according to our validation strategy, both short-term and long-term personalized retrieval is experimented using respectively domains with few queries and other ones with a greater number of queries. Indeed, using few data in the user profile learning process, provides clues on the user intent across few search sessions, while much more data available in the learning process provides a more stable information on the user's general interest.

   Table (2) summarizes the characteristics of our test collection.

Algorithm 1 The test-training based validation strategy

// Test the queries Q^k belonging to the domain C_k
$i = 0$
repeat
  Divide the query domain set into $Q_{training}$ as the training set and Q_{test} as the remaining test set
  //Build a user interest C_i^k on the set $Q_{training}$
  $C_i^k = \{\}$
  for each query q in $Q_{training}$ **do**
    Extract C_{best}: the top 60 terms from relevant documents by applying the *OKAPI* algorithm
    $Merge(C_i^k, C_{best})$
  end for
  Test the Q_{test} queries using C_i^k
  $i++$
until testing all the queries in Q^k
Average the performances over the test query runs Q_{test}

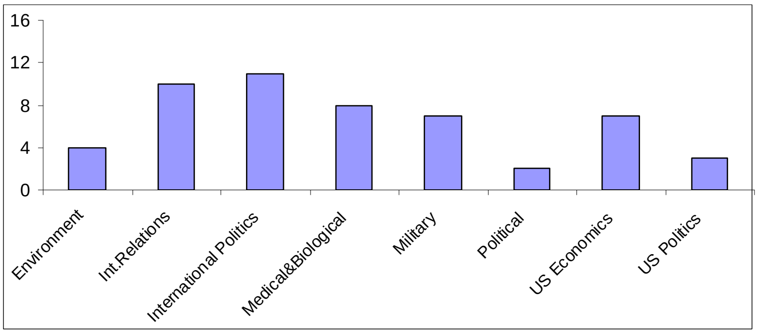

Figure 2: Distribution of query domains

5.2 Performance measures

In order to measure the performances, we used the standard TREC evaluation measures: precision at point 10 ($P@10$) and mean average precision (MAP). $P@10$ is the ratio of relevant documents among the top 10 retrieved documents and the MAP is the mean precision values after each retrieved document. Thus $P@10$ is a high precision oriented measure while MAP score makes some use of recall in its computation. For each query, 1000 top documents are retrieved, we give the average results by means of $P@10$ and MAP over all the queries belonging to each tested domain.

Table 2: Statistics of the data set test

Number of domains	8
Number of documents	3576208
Number of queries	52
Number of distinct terms	589212
Average document length	53,64
Average query length	3,5

6 Experimental results

We attempt to achieve through our experiments three main objectives:

1. Evaluating the effectiveness of our model to enhance retrieval performances by means of personalization. Our aim through the related experiments is to show how well our model is able to exploit evidence from the user interests in order to improve document ranking. In order to achieve this objective, we report below comparative results of two scenarios: with personalization and without personalization according to two different baseline models.
2. Evaluating the effectiveness of our ID based model comparatively to another personalized retrieval model. We attempt to show through the experiments the impact of both using ID as a tool for personalized relevance estimation and exploiting the user interests at the retrieval stage rather than the re-ranking stage.
3. Tuning the aggregation operator according to the relatedness of the user interests. Our objective behind this experimental study is to highlight the importance of an appropriate choice of aggregation operators w.r.t. our decision-making based model.

The student's paired t-test was employed to determine the statistical significance of the results; siginificance at the 5% level is indicated either by $^{\triangle}$ or $^{\triangledown}$ depending on the direction change comparatively to the best baseline result ; no significance is denoted by $^{\diamond}$.

6.1 Personalization vs. no personalization

It is important to determine the performance contribution of each component within our retrieval framework. For this reason, we compared our ID based personalized retrieval model to two baseline models:

1. A Naïve (simple) Bayesian (NB) model that do not deal with the utility measures and the context. Applying the Bayesian joint law on our diagram presented in Figure 1 leads to compute:

$$RSV(Q,D) = \sum_{\theta^s \in \theta} p(q|\theta^s) * \prod_{T_j \in (\tau(Q) \cap \tau(D))} p(\theta_j^s|d)$$

   Our main objective through this comparative evaluation is to highlight the benefit behind adding context to a Bayesian model by means of an ID.

2. The Okapi's model as it performs very well in average for various TREC tasks (Robertson et al., 1992). Our purpose is to highlight how much is effective our model to exploit evidence from the user interests in order to enhance retrieval accuracy.

   In all the experiments the term TF-IDF document weighting formula used is: $Wtd(j,i) = \frac{tf_j * log((n-n_j+0.5)/(n_j+0.5))}{(2*(0.25+0.75*dl_i/avg_{dl})+tf_j)}$, where tf_j is the frequency of term T_j, n_j is the number of documents indexed by T_j, dl_i is the length of document d_i, avg_{dl} is the average document length in the colllection and n is the collection size.

Table (3) presents the average retrieval performance measures $P@10$ and MAP across the 8 tested domains. We can notice that our personalized IR model is effective and achieves

in average statistically significant performance improvements comparatively to both NB's (+84,9%;+60,80%) and Okapi's models (+19,51%;+27,30%). This supports our claim that:
(1) the decision-making inference supported by the use of the utility theory is appropriate to address the personalization task,
(2) introducing domain knowledge at the retrieval level yields substantial improvements over state-of-the art ranking models based solely on evidence issued from the query.

	Okapi's model		**NB's model**		**Our model**	
Domain	*P@10*	*MAP*	*P@10*	*MAP*	*P@10*	*MAP*
Environment	0,47	0,29	0,25	0,17	$0,62^{\triangle}$	$0,29^{\diamond}$
Inter. Relations	0,21	0,06	0,16	0,05	$0,32^{\diamond}$	$0,12^{\triangle}$
Inter. Politics	0,20	0,08	0,11	0,07	$0,24^{\triangle}$	$0,12^{\triangle}$
Medical-Biological	0,38	0,06	0,17	0,03	$0,35^{\triangledown}$	$0,11^{\triangle}$
Military	0,27	0,10	0,25	0,08	$0,37^{\triangle}$	$0,15^{\triangle}$
Political	0,05	0,00	0,00	0,00	$0,20^{\triangle}$	$0,01^{\triangle}$
US Economics	0,28	0,10	0,35	0,12	$0,34^{\triangle}$	$0,12^{\triangle}$
US Politics	0,5	0,06	0,23	0,06	$0,40^{\triangledown}$	$0,07^{\triangle}$
Average	0,29	0,09	0,19	0,07	$0,35^{\triangle}$	$0,12^{\triangle}$
Improvement	**+19,51%**	**+27,30%**	**+84,09%**	**+60,80%**		

Table 3: Comparative retrieval performance with and without personalization

We can also observe that regardless of the large differences between the two baseline performances, the degree of improvement varies significantly from a domain to another; for example, the improvements achieved at (*P*@10, *MAP*) according to the NB and Okapi baseline models for the domain *International Relations* are respectively (+52%;+93%) and (+100%;+108%) and for the domain *US Politics*, the performance results are respectively (-2,10%,+50%) and (71%, +10,5%). This variation could be explained by three main reasons. The first one is related to the variation of the size of the query training sets as illustrated in Figure 2. We expect that when the query training set is small (few queries), there is not sufficient data to learn the "*accurate*" profile leading to better performances. The second reason concerns the variation of query lengths across the domains. Our model may perform differently in average for short vs. long queries. The third reason is related to the performance level of the baseline. Queries with low performance at the baseline might be difficult to be improved and so only slight improvements could be achieved using our model. Effective queries with high performance at the baseline may do not take advantage using extra knowledge from the user interests. Below, experimental investigation is made to clarify the two latter points. More precisely, we focus on the ability of our model to improve short vs. long queries and to improve difficult queries, having worse baseline performances.

6.1.1 Improving short vs. long queries

In order to measure the performance variation according to the query lengths, we first classified the test queries according to their length (expressed by their number of distinct terms) as illustrated in Figure 3. As only the title fields are used for our tests, the queries are relatively short, between 1 and 8, 80% among them having less than 6 terms. We retained only the sets containing at least three queries of the same length in order to achieve reliable conclusions and consequently, did not consider the queries of length 6 (q_{84}), 7 (q_{142}, q_{144}) and 8 (q_{100}). We averaged the performance results over all the queries belonging to the same set. The histograms presented in figure 4 represent the *P@*10 and *MAP* improvements variations over the NB's and Okapi's baseline models for query lengths 1-5.

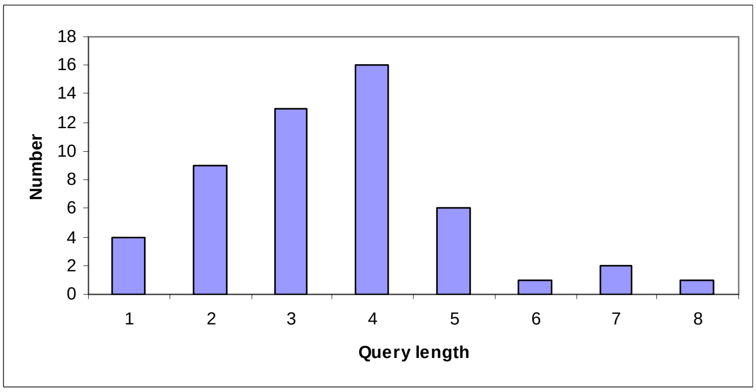

Figure 3: Query length based classifcation

The graphs show that the improvements levels of both *P@*10 and *MAP* are in most of cases positive within the different query lengths. The two baselines corroborate that:
(1) our model performs well either for short queries (length greater than 1) and long queries (length greater than 4),
(2) considering most of the query lengths, *MAP* is better performed than *P@*10.
These results lead us to conclude that our model is not significantly length-sensitive since no significant differences in the performances are observed for short queries comparatively to long ones ($1 \leq query\ length \leq 5$).

6.1.2 Improving difficult queries

We select 28 relatively difficult topics among the test collection on the basis of worse precision results such as $P@10 \leq 0,3$ according to the two baseline models. The reason why we select difficult topics is that we expect that our model to be more useful for enhancing such topics. We divided these topics into three subsets of close sizes, each one is related to an interval value of *P@*10. In order to make our conclusions more reliable, Table (4) presents the improvements achieved comparatively to the Okapi's baseline model that returns better scores in average for initial queries than the NB's model.
It is interesting to notice that the incorporation of the user interests, according to our model, enhances the retrieval effectiveness for difficult queries. However, the improvement scales are significant only for those having baseline precision scores greater than 0. This can be explained by the fact that topics in the same domain can vary significantly and so the user interests built from relevant document training do not sufficiently focus on new query test topics (according to training-test validation startegy) so to improve them; we believe, as claimed in (Bai et al., 2007), that topics in the same domain, especially in large ones like *International Politics* can

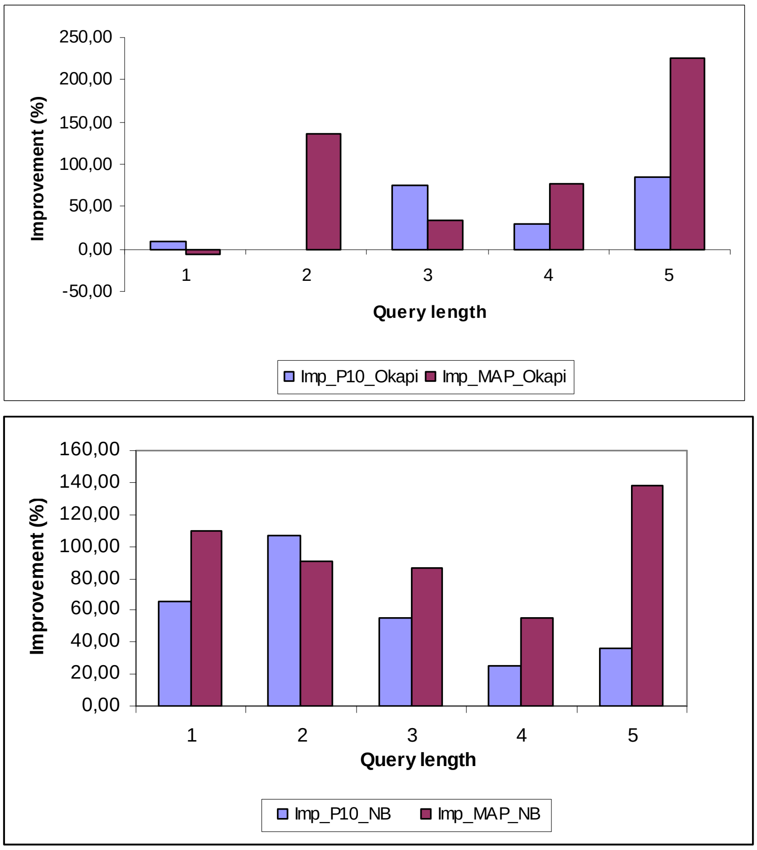

Figure 4: Improvement results for different query lengths

vary greatly and consequently user interests are not able to suggest relevant information to improve the results.

P@10 at the baseline	Number of queries	Improvement
[0 0,1[	11	+3,63%
[0,1 0,2[	10	+110%
[0,2 0,3]	7	+11,11%
	28	41,58%△

Table 4: Improvement results for difficult queries

6.2 Comparing our model to a re-ranking based personalized retrieval model

We compared the effectiveness of our model to a re-ranking based personalized retrieval model derived from the general re-ranking approach proposed in (Speretta and Gauch, 2005). The authors proposed to re-rank the documents by their conceptual similarity to produce their conceptual rank; The final rank of document profile D_i is obtained using the formula:

$$FinalRank(D_i) = \alpha * ConceptualRank(D_i) + (1-\alpha) * InitialRank(D_i)$$

where document profile D_i is the set of concepts issued from mapping document D_i on a reference ontology, *ConceptualRank(D_i)* is the document rank obtained by computing the similarity between document profile D_i and the user profile $User_k$ using the similarity function: $sim(User_k, D_i) = \sum_{l=1}^{m} wt(k,l) + wt(i,l)$, where $wt(k,l)$ is the weight of the concept l in user profile $User_k$, $wt(i,l)$ is the weight of concept l in document profile D_i, *InitialRank* is the initial document rank given by the search engine; in our case, the initial rank of a document is given by the Okapi's search engine. α is a constant having a value between 0 and 1; When $\alpha = 0$, conceptual rank is equivalent to the original rank assigned by the Okapi search engine. If $\alpha = 1$ the initial search engine ranking is ignored and pure conceptual rank is considered. The conceptual and search engine based rankings can be combined in different proportions by varying the value of α.

Before discussing the results, we outline that in order to achieve accurate comparative evaluation according to our framework, the user profile is still be characterised by a set of distinct basic terms (not concepts) built via a simulation based on the Okapi algorithm (the same that those used in our previous experiments). Indeed, our purpose is to compare the performance results obtained with two personalization approaches: our ID based retrieval model and a re-ranking post-retrieval strategy. The impact of the semantic (concept based) vs. the flat term-based representation of the user profile is not the focus of our experimental study. In our experiments, we choose $\alpha = 0.5$ rather than $\alpha = 1$, as claimed by the authors in (Speretta and Gauch, 2005) because it is the value giving the best results considering various parameters: our inverted file built with our indexing method, our strategy for building the user interests, our term based representation of the user profiles.

Table 5 shows the results of the eight (8) runs on the TREC domains. Figure 5 shows respectively the *P*@10 and *MAP* improvements achieved by means of our model comparatively to the re-ranking baseline model for each run. We notice that our model achieves significant positive improvement results, leading to a better average performance. The ID based retrieval model achieves a *P*@10 score (resp. a *MAP* score) that was on average $17,6\%$ (resp. $7,6\%$) higher than the re-ranking post-retrieval model. Regardless of the *Political* run containing few (2) queries, this increase is between $26,34\%$ and $66,67\%$ using the *P*@10 measure and between $28,94\%$ and $64,71\%$ using the *MAP* measure. We also notice that under both the *P*@10 and *MAP* measures, the re-ranking based model achieves better performances (negative bars) than the ID based model for *International Politics* and *US Politics* runs. This result can be explained by the wideness of these domains. Indeed, in the case of wide domains, the initial ranking, exploited *a posteriori* by this model, provides additional clues to better fit the user profile, while the ID based model lacks of such evidence as it performs at the retrieval stage. However, this improvement remains generally dependent on the quality of the initial ranking, that what avoids our model. It is left to future work to investigate whether adjustments to our probability calculations that take into account information about the domain characteristics, be beneficial at the retrieval level. It is a potential research direction that we currently explore by means of semantic representations of the user profile, driven by ontologies *(Daoud et al., 2009)*.

	Re-ranking model		Our model	
Domain	*P@10*	*MAP*	*P@10*	*MAP*
Environment	0,37	0,24	$0,62^{\triangle}$	$0,29^{\triangle}$
Inter. Relations	0,25	0,1	$0,32^{\triangle}$	$0,12^{\triangle}$
Inter. Politics	0,34	0,13	$0,24^{\triangledown}$	$0,12^{\triangledown}$
Medical-Biological	0,21	0,06	$0,35^{\triangle}$	$0,11^{\triangle}$
Military	0,38	0,13	$0,37^{\diamond}$	$0,15^{\triangle}$
Political	0,05	0,01	$0,20^{\triangle}$	$0,01^{\diamond}$
US Economics	0,27	0,14	$0,34^{\triangle}$	$0,12^{\triangledown}$
US Politics	0,53	0,11	$0,34^{\triangledown}$	$0,12^{\triangledown}$
Average	0,30	0,11	$0,35^{\triangle}$	$0,12^{\triangle}$
Improvement			**+17,89%**	**+7,6%**

Table 5: Comparative retrieval performances of our model vs. a re-ranking based model

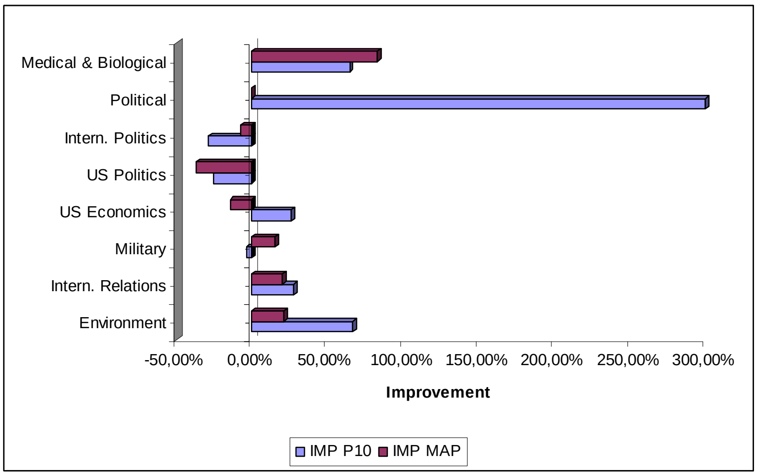

Figure 5: Comparison of improvement results for different domains

6.3 Impact of the aggregation operator

At this level, we focus on the choice of a suitable aggregation operator to accurately combine the evidence extracted from each of the various user interests present at the broader context of his information seeking. Our intuition behind these experiments, is that the impact of the aggregation operator depends mainly on the dependency vs. independency of the topics of interests as stated in hypothesis 1 and hypothesis 2, previously assumed in paragraph 4.2.2. In order to detect dependencies among domains, we choose for our experiments two pairs of domains:

1. *a priori* dependent domains: $(dom_1, dom_2) = argmax_{(dom_i, dom_j) \forall i \neq j} Sim(C_i, C_j)$
2. *a priori* independent domains: $(dom_1, dom_2) = argmin_{(dom_i, dom_j) \forall i \neq j} Sim(C_i, C_j)$

where C_i (resp. C_j) is the user interest extracted from all the queries of dom_i (resp. dom_j), *Sim* is the cosinus similarity measure.
The computation of similarities on our test collection considering our strategy of building simulated user interests, reveals that only the pair *(Environment, Military)* returns a not null

similarity $(0,28)$. We choose then to exploit this pair and another randomly pair *(Environment, US Politics)*. We expressed respectively Ψ_{ind} and Ψ_{dep} specified in paragraph 4.2.2., using an equivalent full ordering function that allows us to compute the basic performance metrics as follows:

- Ψ_{ind}: documents are ranked according to $Max(RSV_{C_1}(D_i),\ldots,RSV_{C_p}(D_i))$
- Ψ_{dep} : documents are ranked according to $Sum(RSV_{C_1}(D_i),\ldots,RSV_{C_p}(D_i))$

We observe, through Table (6), that in the case of *a priori* related domains, both *Max* and *Sum* operators perform equally.

	Ψ_{ind}		Ψ_{dep}	
Domains	*P@10*	*MAP*	*P@10*	*MAP*
Environment	0,62	0,29	0,62	0,29
Military	0,37	0,15	0,37	0,15

Table 6: Impact of the aggregation operator on the retrieval performances: case of a priori related domains (a)

	Ψ_{ind}		Ψ_{dep}	
Domains	*P@10*	*MAP*	*P@10*	*MAP*
Environment	0,62	0,29	0,65	0,29
US Politics	0,4	0,07	0,63	0, 11

Table 7: Impact of the aggregation operator on the retrieval performances: case of a priori unrelated domains (b)

The positive point to be retained is that the utility aggregation does not decrease the performances at all. As the results are equivalent to those obtained considering only one general user interest, we can conclude at this level that:

- the domains are not really content dependent and/or,
- the aggregation operators choosen for our experiments are not suitable to combine the evidence extracted from the different user interests.

We investigated this question in the following, in order to better understand this observation.

We notice, in Table (7), that in the case of *a priori* unrelated domains, it seems that the *Sum* operator performs better for especially the domain *US Politics*. We anlyzed then the results per query for the domain *US Politics*. This analysis reveals that the improvement is only due to the query q_{145} for which *P*@10 and *MAP* have been boosted respectively from (0,2;0,04) to (0,9;0,15). An analysis of the topic description of q_{145} from the domain *US Politics* and q_{78} from the domain *Environment*, presented below, highlights the reason:

```
Number: 078
Domain: Environment
title Topic: Greenpeace
desc Description: Document will report activity by Greenpeace to carry out
their environmental protection goals
smry Summary: Document will report activity by Greenpeace to carry out
their environmental protection goals
Concept(s):
1.  Greenpeace, environment, group, activist
2.  protest, disrupt, block, harass, scuttle, trespass, confront
3.  anti-nuclear, uranium, radioactive, missile
```

```
Number: 145
dom Domain: U.S. Politics
title Topic: Influence of the "Pro-Israel Lobby"
desc Description: Document will describe how, and how effectively,
the so-called "pro-Israel lobby" operates in the United States.
smry Summary: Document will describe attempts by the so-called
"pro-Israel lobby" to influence United States policy
Concept(s):
1. zionism, American Jews, Jewish community, U.S. Jewish leaders
2. aid to Israel, military assistance, campaign contribution
3. U.S. arms sales to Egypt, Jordan, Saudi Arabia, or Kuwait
4. U.S. supporters of Israel, pro-Israel congressman or senator,
pro-Israel lobbyist, Jewish lobby
```

The query q_{145} is improved by exploiting relevant terms suggested by the user interest built from the domain *Environment* and consequently the aggregation based on the *Sum* operator that favours the average aggregation, is more effective than the aggregation based on the *Max* operator that favours the strict one. The results lead us to confirm our intuition on the usefulness of tuning the aggregation operator on the basis of the relatedness between the user interests but the challenge in future, is to detect suitable indicators for measuring such relatedness. Our experiments allow us to conclude that basic similarities between the user interests built using the related domains are not sufficient; statistical term distributions in-domain and intra-domain shall be computed in order to better fit the subtopics related to different queries belonging to the same or the different domains of interest.

7 CONCLUSION AND FUTURE WORK

The research contribution presented in this paper offers a theoretical support that consolidates personalized IR applicable to a wide set of IR applications in various domains, such as library, medicine or legal IR. The novelty of our approach concerns mainly the use of decision-making theory in the information personalization process. The proposed model is based on an extension of Bayesian networks, namely influence diagrams, allowing the computation of document utility as a measure of the usefulness of a document for a specific user having a broad variety of interests. Query evaluation is viewed as an inference process involved through a diagram as in Bayesian networks, via the computation of *a posteriori* probabilities as evidence measures attached to various kinds of information (terms, documents and user interests) in order to compute accurate global document utility values. Furthermore, the evidence extracted from the various topics of interests is aggregated according to basic hypotheses about the relatedness of the user interests contents.

A framework evaluation based on a specific TREC sub-collection is proposed. The experimental results illustrate that the model is successful at selecting more relevant documents according to the user's topics of interest comparatively to both the naive Bayesian model and Okapi's model. Comparative experimental evaluation with a post-retrieval personalized retrieval model, reveals that our model is more effective on the average, but we should investigate the impact of probability calculations on the search accuracy, particularly in the case of large domains.

Furthermore, The aggregation operators are appropriate tools enabling suitable pooling of the evidence extracted from the user profile but should be much more explored in order to achieve more reliable conclusions.

Future research will focus first on the refinement of the decision-theoretical framework from the quantitative aspect. We plan to explore various probability, utility formula and aggregation functions on the basis of evidence surronding the IR process. The second investigation will deal with the empirical study. The proposed model will be tested with multiple users in an empirical setting in order to gauge its faisablility in real life IR applications.

REFERENCES

Agichtein, E., Brill, E., and Dumais, S. T. (2006). Improving web search ranking by incorporating user behavior information. In *Proceedings of the* 29^{th} *International SIGIR conference on research and development in information retrieval*, pages 19–26.

Arampatzis, A., Beney, J., Koster, C., and Weide, T. (2001). Incrementality, half-life, and threshold optimization for adaptive document filtering. In *Proceedings of the* 9^{th} *Text REtrivel Conference (TREC-9)*, pages 589–602. NIST Special Publication.

Bai, J., Nie, J., Bouchard, H., and Cao, G. (2007). Using query contexts in information retrieval. In *Proceedings of the 31 st annual internationsl ACM SIGIR Conference on Research and development in Information retrieval*, pages 15–22.

Beitzel, S. M., Jensen, E., Lewis, D., Chowdhury, A., and Frieder, O. (2007). Automatic classification of web queries using very large unlabeled query logs. *ACM Transactions on Information Systems*, 5(2):9.

Borlund, P. (2003). The concept of relevance in ir. *Journal of the American Society for Information Science and Technology*, 54(10):913–925.

Cheverst, K., Davies, N., Mitchell, K., Friday, A., and Efstratiou, C. (2000). Developing a context-aware electronic tourist guide: some issues and experiences. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 17–24, New York, NY, USA. ACM.

Cleverdon, C. (1967). The cranfield test on index language devices. In *Aslib*, pages 173–194.

Crestani, F. and Ruthven, I. (2007). Introduction to special issue on contextual information retrieval systems. *Information retrieval (IR)*, 10:829–847.

Daoud, M., Tamine, L., Boughanem, M., and B.Chebaro (2009). A session based personalized search using an ontological user profile. In *ACM Symposium on Applied Computing (SAC), Haiwai (USA), 08/03/2009-12/03/2009*, pages 1031–1035, http://www.acm.org/. ACM.

Daoud, M., Tamine-Lechani, L., Boughanem, M., and B.Chebaro (2008). Learning user interests for session-based personalized search. In Borlund, Schneider, Lalmas, and Tombros, editors, *ACM Information Interaction in context (IIiX), London, 14/10/2008-17/10/2008*, pages 57–64, http://www.acm.org/. ACM.

Davies, N., Mitchell, K., Cheverest, K., and Blair, G. (1998). Developing a context sensitive tourist guide. In *First Workshop on Human Computer Interaction with Mobile Devices*. GIST Technical Report G98-1.

De Mey, M. (1977). The cognitive viewpoint: Its development and its scope. *International Workshop on the cognitive viewpoint*, pages 285–295.

Dervin, B. and Nilan, M. (1986). Information needs and uses. *ARIST, William, M.E. Eds*, pages 3–33.

Ding, C. and Patra, J. C. (2007). User modeling for personalized web search with self-organizing map. *Journal of American Society in Information Science and Technology*, 58(4):494–507.

Fargier, H. and Perny, P. (2003). A characterisation of generalised concordance rules in multicriteria decision making. *In International Journal of Intelligent Systems (IJIS)*, 18:751–774.

Gowan, J. M. (2003). *A multiple model approach to personalised information access*. Master Thesis in computer science, Faculty of science, University College Dublin.

Gravano, L., V., V. H., and Lichtenstein, R. (2003). Categorizing web queries according to geographical locality. In *International ACM Conference on Information Knowledge and Management (CIKM)*, pages 325–333.

Harman, D. (1988). Towards interactive query expansion. In *Proceedings of the* 11^{th} *annual international ACM SIGIR conference on Research and development in information retrieval*, pages 321–331.

Hattori, S., Tezuka, T., and Tanaka, K. (2007). Context-aware query refinement for mobile web search. In *Proceedings of the 2007 International Symposium on Applications and the Internet Workshops*, page 15, Washington, DC, USA. IEEE Computer Society.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction:elements of a cognitive theory. *Journal of documentation*, 52(1):3–50.

Ingwersen, P. and Jarvelin, K. (2005). *The Turn: Integration of information seeking and information retrieval in context*. Springer.

Jeh, G. and Widom, J. (2002). Scaling personalized web search. In *Proceedings of the* 12^{th} *international conference on World Wide Web*, pages 271–279, New York, NY, USA. ACM Press.

Jensen, F. (2001). *Bayesian networks, decision graphs*. Springer.

Joachims, T., Granka, L., Hembrooke, H., Radlinski, F., and Gay, G. (2007). Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information systems*, 25(2):7.

Jose, J. and Rijsbergen, C. v. (2005). Workshop on information retrieval in context: Report. In *Proceedings of the Workshop on Information Retrieval in Context, LNCS, Springer Eds*.

Kang, I. and Kim, G. (2003). Query type classification for web document retrieval. In *Proceedings of the* 26_{th} *International SIGIR conference on research and development in information retrieval*, pages 64–71.

Kelly, D. and Belkin, N. (2004). Display time as implicit feedback: Understanding task effecys. In *Proceedings of the* 27^{th} *annual international ACM SIGIR conference on Research and development in information retrieval*, pages 377–384.

Kelly, D. and Fu, X. (2007). Eliciting better information need descriptions from users of information search systems. *Information Processing and Management*, 43(1):30–46.

Lee, J., Hu, X., and Downie, J. (2005). Qa websites: Rich research resources for contextualizing information retrieval behaviors. In *Proceedings of the* 28^{th} *International SIGIR conference on research and development in information retrieval, Workshop on information retrieval in context*, pages 33–366.

Leuski, A. (2005). Context features in email archives. In *Proceedings of the* 28^{th} *International SIGIR conference on research and development in information retrieval, Workshop on information retrieval in context*, pages 54–56.

Lieberman, H. (1995). Letizia: An agent that assists web browsing. In *Proceedings of the* $14^{t}h$ *International Joint Conference On Artificial Intelligence*, Montreal, Canada.

Lin, J. and Demner-Fushman, D. (2005). Representation of information needs and the elements of context: a case study in the domain of clinical medicine. In *Proceedings of the* 28^{th} *International SIGIR conference on research and development in information retrieval, Workshop on information retrieval in context*, pages 51–53.

Liu, F. and Yu, C. (2004). Personalized web search for improving retrieval effectiveness. *IEEE Transactions on knowledge and data engineering*, 16(1):28–40.

Magennis, C. and Rijsbergen, J. v. (1997). The potential and actual effectiveness of interactive query expansion. In *Proceedings of the* 20^{th} *annual international ACM SIGIR conference on Research and development in information retrieval*, pages 324–332.

Micarelli, A. and Sicarrone, F. (2004). Anatomy and empirical evaluation of an daptive web-based information filtering system. *User Modeling and User-Adapted Interaction*, 14(2-3):159–200.

Mitra, M., Singhal, A., and Buckley, C. (1998). Improving automatic query expansion. In *Proceedings of the* 21^{st} *annual international ACM SIGIR conference on Research and development in information retrieval*, pages 206–214.

Mobasher, B. (2007). *Data mining for Web personalization*. Brusilovsky, P. Kobsa, A. Nejdl W., Lecture Notes in Computer Science. Springer Verlag, Berlin Heidelberg, NewYork.

Mobasher, B., Dai, H., Luo, T., Sun, Y., and Zhu, J. (2000). Integrating web usage and content mining for more effective personalization. In *EC-Web*, pages 165–176.

Mostafa, J., Mukhopadhyay, S., and Palakal, M. (2003). Simulation studies of different dimensions of users' interests and their impact on user modeling and information filtering. *Information Retrieval*, 6(2):199–223.

Muresan, G., Smith, C. L., Cole, M., Liu, L., and Belkin, N. J. (2006). Detecting document genre for personalization of information retrieval. In *Proceedings of the 39th Annual Hawaii international Conference on System Sciences*, pages 50c– 50c. HICSS, IEEE Computer Society, Washington.

Nottelman, H. and Fuhr, N. (2003). From retrieval status values to probabilities of relevance for advanced information retrieval applications. In *Proceedings of 25th European Conference on IR Research, ECIR*, pages 235–250. Springer Lecture Notes in Computer Science.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible Inference*. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA. isbn: 0-934613-73-7.

Robertson, S. E., Walker, S., Hancock-Beaulieu, M., Gull, A., and Lau, M. (1992). Okapi at trec 3. In *Text REtrieval Conference*, pages 21–30.

Ruthven, I. Lalmas, M. (2003). A survey on the use of relevance feedback for information access systems. *Kowledge engeneering review*, 18(2):95–145.

Ryan, N., Pascoe, J., and D.Morse (1997). *Enhanced Reality Fieldwork: the Context-Aware Archaeological Assistant*. Gaffney, V., van Leusen, M., Exxon, S. (eds.) Computer Applications in Archeology.

Schilit, B., Adams, N., and Want, R. (1994). Context-aware computing applications. In *Proceedings of the Workshop on Mobile Computing Systems and Applications*, pages 85–90. IEEE Computer Society, Santa Cruz, CA.

Shachter, R. (1988). Probabilistic inference and influence diagrams. *Operating Research*, 36(4):589–604.

Shamber, L. (1994). Relevance and information behaviour. *ARIST, William, M.E. Eds*, pages 3–48.

Shen, X., Tan, B., and Zhai, C. (2005). Context-sensitive information retrieval using implicit feedback. In *In Proceedings of the 29th annual internationsl ACM SIGIR Conference on Research and development in Information retrieval*, pages 43–50.

Sieg, A., Mobasher, B., and Burke, R. (2004). User's information context: Integrating user profiles and concept hierarchies. In *Proceedings of the 2004 Meeting of the International Federation of Classification Societies*, number 1, pages 28–40.

Speretta, M. and Gauch, S. (2005). Personalized search based on user search histories. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, pages 622–628.

Tamine, L., Boughanem, M., and Daoud, M. (2009). Evaluation of contextual information retrieval: overview of issues and research. *Knowledge and Information Systems (Kais), to appear*.

Tamine, L., Chrisment, C., and Boughanem, M. (2003). Multiple query evaluation based on an enhanced genetic algorithm. *Information Processing and Management (IPM)*, 39(2):215–231.

Tamine-Lechani, L., Boubekeur, F., and Boughanem, M. (2007). On using graphical models for supporting context-aware information retrieval. In *International Conference on the Theory of Information Retrieval (ICTIR), Budapest (Hungary), 18/10/2007-20/10/2007*, pages 213–222. Foundation for Information Society.

Tao, Y., Mamoulis, N., and Papadias, D. (2003). Validity information retrieval for spatio-temporal queries. In *SSTD*.

Tebri, H., Boughanem, M., Chrisment, C., and Tmar, M. (2005). Incremental profile learning based on a reinforcement method . In *SAC'2005- 20th ACM Symposium on Applied Computing , Santa Fe, New Mexico, USA, 13/03/05-17/03/05*, pages 1096–1101. ACM.

Teevan, J. and Dumais, S. (2005). Personalizing search via automated analysis of interests and activities. In *Proceedings of the* 28_{th} *International SIGIR conference on research and development in information retrieval*, pages 449–456.

Turtle, H. and Croft, W. (1990). Inference networks for document retrieval. In *Proceedings of the* 13^{th} *International Conference on Research and Development in Information Retrieval (SIGIR)*, pages 1–24.

Watters, C. and Amoudi, G. (2003). Geosearcher: location-based ranking of search engine results. *Journal of American Society on Information Science and Technology*, 54(2):140–151.

Webb, G.I. Pazzani, M. B. D. (2001). Machine learning for user modeling. *User Modeling and User-Adapted Interaction*, 11(1-2):19–29.

Westerveld, T., Kraaij, W., and Hiemstra, D. (2001). Retrieving web pages using content, links, urls and anchors. In *In: Proceedings of the Tenth Text REtrieval Conference (TREC 2001)*, pages 663–672.

Yau, S., Huan, L., Huang, D., and Yao, Y. (2003). Situation-aware personalized information retrieval for mobile internet. In *Proceedings of the* 27^{th} *Annual International Computer Software and Applications Conference (COMPSAC)*.

Zemirli, N., Tamine, L., and Boughanem, M. (2007). A personalized retrieval model based on influence diagrams. In *Proceedings of Context Information Retrieval Workshop*. Roskilde University, Denmark, 20-AUG-07-24-AUG-07, Doan Bich-Liên, Jose Joemon, Melucci Massimo (Eds.).

Zhang, Y. and Callan, J. (2001). Maximum likelihood estimation for filtering thresholds. In *Proceedings of the* 24^{th} *annual international ACM SIGIR conference on Research and development in information retrieval*, pages 294–302.